# Sampling and Reconstructing User Experience

*Panos Markopoulos, Eindhoven University of Technology, The Netherlands*

*Vassilis-Javed Khan, NHTV Breda University of Applied Sciences, The Netherlands*

## ABSTRACT

*The Experience Sampling and Reconstruction (ESRM) method is a research method suitable for user studies conducted in situ that is needed for the design and evaluation of ambient intelligence technologies. ESRM is a diary method supported by a distributed application, Reconexp, which runs on a mobile device and a website, enabling surveying user attitudes, experiences, and requirements in field studies. ESRM combines aspects of the Experience Sampling Method and the Day Reconstruction Method aiming to reduce data loss, improve data quality, and reduce burden put upon participants. The authors present a case study of using this method in the context of a study of communication needs of working parents with young children. Requirements for future developments of the tool and the method are discussed.*

*Keywords:* *Awareness Systems, Day Reconstruction Method, Experience Sampling Method, Family Communication, In Situ Requirements Elicitation, Mobile Survey, User-Centered Design*

## INTRODUCTION

The current trend towards pervasive and context sensitive applications where information and computational technology are embedded in our social and physical environments presents substantial methodological challenges for researchers, designers, or technologists, wishing to design, analyze, or evaluate, corresponding user experiences. Available research methods have been shaped in past decades to support the design and evaluation of the cognitive ergonomics of task-oriented interaction, usually contained within a contained time span. Extending such methods to study user experiences as these occur in situ, unfolding over days or weeks, capturing social interactions between several people and diverse environmental and technical contingencies, requires a substantial scaling up the data sampling in terms of frequency, duration, and the richness of records made.

The objectives of system evaluation have also changed significantly. Transcending usability, evaluations of applications and services that are mobile and often context sensitive, typically examine higher level aspects of user experiences and user needs relating to persuasion, fun, engagement, trust, etc. Contextualized methods of data collection need to support the reporting of attitudes, opinions, or appraisals, close to the moment that a particular experience occurs and in the context where events and activities unfold. Such surveying of user attitudes can

DOI: 10.4018/jhcr.2011070104

occur repeatedly over time, allowing the study of behaviors and experiences over medium or long periods of time, even enabling researchers to examine patterns of use over time.

A well established method that addresses these requirements to a large extent is the *diary method* whereby informants are asked to keep a journal or a log, where they record events, activities and experiences regularly over a specified period of time (Rieman, 1993). In traditional diary studies informants record data, usually in writing, but often combining or even replacing written records with other recording media, see for example Carter and Mankoff (2005).

In diary studies, the initiative for capturing information is left up to the study participants who have to remember and take the initiative to report in their diaries. This may be detrimental to the quality of the data collected for several reasons. Participants may forget to enter information in diaries, or entries may be made at moments that they have the time and appetite to do so, rather than the ones of interest to the researcher. This can lead to loss of data and systematic response biases.

For these reasons, the Experience Sampling Method (ESM) is gaining ground in human-computer interaction studies for understanding human behavior to design better products and services and for studying use in the field. The ESM is a quasi-naturalistic method that involves signaling questions at informants repeatedly throughout the sampling period. In its original form (Csiekszentimihalyi & Larson, 1987) Experience Sampling method required participants to carry a pager or any another notification device through which they would be reminded to fill in a set of questions in a paper diary . With developments in handheld computing, this method has become computerized and a variety of tools have been developed for handheld computing devices to support it (Barret & Barret, 2001). Participants are typically requested to carry a dedicated handheld device for the whole study period through which a predefined question-asking protocol is applied.

The ESM method is gaining in popularity in the field of human-computer interaction. Consolvo and Walker (2003) have used the ESM for evaluating an Intel Research system called Personal Server. Hudson et al. (2002) have used the ESM to explore attitudes about availability of managers at IBM Research. Froehlich et al. (2006) used it to investigate the relationship between explicit place ratings and implicit aspects of travel such as visit frequency. The list is longer and growing rapidly as this field turns its attention towards the design and evaluation of mobile and context sensitive applications.

## CHALLENGES AND PITFALLS OF EXPERIENCE SAMPLING

Although very useful in prompting the reporting of subjective experiences over time and in context, ESM also has shortcomings such as interrupting the subject at inappropriate moments, the onus of repeatedly answering the same or similar questions, the difficulty of entering self-report data in inconvenient social and physical contexts. The main consequence of these is loss of data: participants ignoring the alert and refraining from providing the requested self-report. Loss of data seems to be a major problem. Froehlich et al. (2006) report completion rate of 80.5% similar to Consolvo and Walker (2003) who report an 80% completion rate (on average 56 out of 70) with as low as 28.5% (20 out of 70). Even worse, these numbers are silent regarding the significance of the data lost. It is reasonable to assume that data loss occurs when people are busy or engaged in social or professional activities. Depending on the goals of the investigator, these might be precisely the situations that researchers are interested in studying.

Another problem with the traditional form of experience sampling when applied to the study of specific user experiences has to do with the density and coverage of the sampling. Contrary to studies of emotions and

moods, as was the original application of the method, usage of applications and services may be sporadic and tied to specific contexts. A homogeneous sampling is then inappropriate and researchers need to sample more frequently when specific activities take place and less frequently otherwise, or to adapt the frequency to specific contexts. This requires the use of more sophisticated sampling schemes that can be preprogrammed or even context sensitive. Thus, Intille et al. (2003) have developed software that enables researchers to acquire feedback from participants only in particular situations that are detected by sensors connected to a handheld computing device. Froehlich et al. (2007) developed MyExperience, a system for capturing both objective and subjective in situ data on mobile computing activities. A common characteristic of these works is that they aim to optimize the choice of when to prompt informants with a question. This choice can be based on previous answers of a participant or on inferences made regarding their activity based on context sensing, see for example, Kapoor and Horvitz (2008). Results on this direction are positive but the fundamental limitation of ESM noted above resulting in loss of data remains.

Regardless of the possibility to vary and adapt the timing and frequency of sampling ESM is inherently expensive; it puts a high burden on participants who may have to interrupt own activities and spend effort in situ reporting on their experiences. Further, despite the technical developments described above, it can only provide limited information about uncommon or brief events, which are rarely sampled (Kahneman et al., 2004).

These problems lead to loss of data, inaccurate reporting and nuisance to participants. Current research in this field is concerned with developing methodological innovations and corresponding tools to remedy these shortcomings. On the other hand, the unique advantage of ESM is its ability to capture daily life as it is directly perceived from one moment to the next (Froehlich et al., 2007), providing a rich set of data to researchers.

## THE DAY RECONSTRUCTION METHOD

An alternative to ESM, proposed by Kahneman et al. (2004) is the Day Reconstruction Method (DRM), which was designed to assess how people experience their various activities and settings of their lives. Participants in this case are asked to record a detailed diary of activities and events during one day. These do not relate directly to the focus of inquiry of the researcher, which is not disclosed to them at this point, but are meant as a memory aid, a kind of scaffolding, to allow informants to recall and reconstruct the experiences and feelings of the last day during a follow up interview the day after. This is an in-depth semi-structured interview, during which the researcher probes regarding experiences and feelings that the investigation aims to explore. Kahneman produced strong evidence regarding the efficacy of this method; however by its nature, DRM is more appropriate for short studies. Its efficacy for providing rich and contextualized accounts of user emotions in the last 24 hours is achieved by means of an elaborate interview which is not meant to be carried out repeatedly in a study and is practically difficult to repeat over longer sampling periods. Field studies in the domain of ambient intelligence typically exceed two weeks in duration, reaching sometimes even half a year. For such cases, and for studying patterns in the data over time, DRM can help understand only a small fraction of the activities and experiences of informants, missing out a lot of information regarding the context in which it takes place.

A combination of ESM and DRM has the potential to compensate for their complementary weaknesses. Such a combination is the Experience Sampling and Reconstruction Method (ESRM) introduced below.

### Experience Sampling and Reconstruction Method

ESRM is a combination of Experience Sampling and the Day Reconstruction methods described

in the previous section. Following this hybrid method participants follow the same procedure as with ESM through which a partially complete record of their experiences is created (given the data loss issues discussed above). At regular intervals (e.g., daily) participants are required to reconstruct, elaborate and even reflect on the reported experiences using the partially complete ESM log as a scaffold. The experience can be reconstructed in a manner similar to DRM by completing gaps in the data collection of the day and partly by elaborating and reflecting on this recent data. Because this part of the self-reporting does not have to be done in situ and can be postponed slightly, it can be done using a desktop device that supports more efficient visualization and data entry than handheld devices. Crucially, and thank to system support, this stage is still lightweight enough that it can be repeated daily. The queries which are missed during the sampling day can then be recovered in this way.

The motivation for the reconstruction is twofold. The first motive is to retrieve data which are lost, e.g., because participants are prompted to report at moments inconvenient for them, e.g., when in a meeting, or when driving, etc. The second motive is to provide a way for participants to annotate the samples taken during the day. By its nature, the ESM requires that responses solicited from informants should be brief to reduce disruption and encourage frequent reporting. This though can come at the expense of capturing rich and detailed information about the experiences studied. The reconstruction step allows for more comprehensive reporting and more reflective comments to be provided by informants than by the traditional ESM protocol alone.

In contrast to the classical DRM the purpose of the sampling is divulged to participants when they construct the log of experiences. Further, the sampling and reconstruction activities may take place for long durations as with ESM. Contrary to ESM, informants are allowed to report data post-hoc and even elaborate and modify their earlier responses. This, it is hoped, should address potential loss of data of the ESM and recollection problems that may occur with DRM.

## ESRM PROCEDURE

The procedure is described below from the perspective of the participant; the required initialization of the application by the researcher is self-evident so it is omitted.

The exact procedure for an informant is as follows (Figure 1):

- Personalization of experience sampling protocol on desktop device.
- Combined ESM and DRM data collection.
- Debriefing interview.

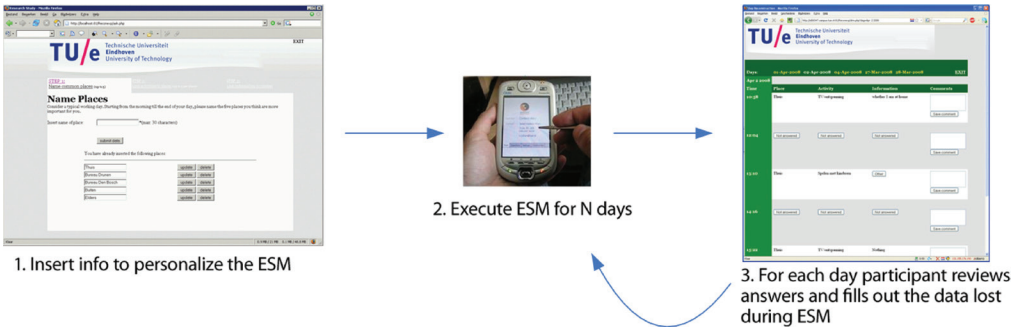
These steps are discussed in detail below.

### ESRM Procedure: Personalization Step

The purpose of having a personalization step is to subsequently minimize the time and effort needed by participants to respond to the mobile device. Reducing the effort required is expected to help prevent data loss but also encourage accurate reporting.

Personalization can mean a few things: adjusting the timing of the sampling procedure or personalizing defaults and choice items offered during the experience sampling part. The personalization step also gives insight into the differences of self-report obtained out of context prior to the sampling period and the data obtained through experience sampling. This comparison can be interesting in itself as it relates participants' expectations with what actually happens and is experienced. Further, the information obtained is used to populate list-boxes offering choices to participants at sampling time rather than requiring text entry. Personalization of the timing for the sampling events can help, e.g., to limit sampling to waking hours, to avoid times where it is known a priori that the participant is inhibited from answering

*Figure 1. The steps of the ESRM method. First, participants insert information to personalize the experience sampling step. Then the experience sampling is executed using a hand-held device and for each experience sampling day participants are asked to access the web application to review their answer and fill out the data which were lost during the sampling day.*




questions, or even to enable dense sampling at times of most interest for the investigator.

## ESRM Procedure: Sampling and Reconstruction Step

During the day the device prompts participants to enter information as in standard experience sampling. Prompting can be programmatically controlled to occur in regular intervals, at random moments or when some conditions regarding the context and the informant activity have been specified.

The informant can respond by selecting between choices of items describing his activity, context or emotions, or even by free text entry to answer more open questions.

The information entered on the handheld device is stored on the online database and is available for retrieval and review directly.

The reconstruction step should happen as close as possible to the collection of data through experience sampling, e.g., within 24 hours. It requires the visualization of the experience sampling logs, the ability to edit them and provide extra information. The interaction requirements for the tool support are different than those applicable for experience sampling: whereas mobility and speed of entry of some brief information is the priority during experience sampling, it is now required to have a good visualization, and efficient ways of editing and inputting text, e.g., using a desktop computer. Of course, one could also allow revision and editing of answers using a small handheld device also for the reconstruction, but this could be at the expense of obtaining richer and more extensive descriptions from informants.

Appropriate visualization of earlier answers can help informants reconstruct their experiences and provide richer descriptions about them. Also important, such visualization can help researchers track the progress of the study, opening up the possibility to modify the sampling protocol while the study unfolds. Researchers can, for example, provide additional incentives or further instructions if they notice that a particular participant is not responding to the daily queries. It also allows researchers to prepare questions for debriefing interviews while the sampling is still unfolding.

## ESRM PROCEDURE: DEBRIEFING INTERVIEW STEP

During the debriefing interview participants are asked to reflect upon their opinions to the queries posed during the sampling period, to solicit in depth explanations and reflection. This step becomes even more useful if the logs of answers are reviewed before approaching each

participant. For example, researchers might spot in the log a pattern in the way a participant had answered to a particular question. Based on such an observation the researcher has a unique opportunity in discussing the pattern in detail with the participant. Moreover, the researcher conducting the interview can go through the logs together with each participant and let participants give further explanations of the underlying reasons behind the answers.

## Implementation: The RECONEXP Tool

The "Reconexp" ("reconstructing experience") tool was developed to support the ESRM method. It is a distributed application partly running on a Smartphone (from now on mentioned as "device") and partly on a website. This section describes how Reconexp embodies the characteristics of the hybrid method and how we used it for the purposes of our research.

Currently the application has been deployed on QTek 9090 and HTC Touch P3450 smartphones. QTek runs Windows Mobile 2003 Second Edition while HTC runs Windows Mobile 6 Professional. The Reconexp application has been programmed using Microsoft's .NET Compact Framework in C#, and OpenNetCF libraries for controlling the WiFi adapter of the device. Participant's data are managed using the MS SQL Server CE. The replication features of Microsoft's SQL Server are used for merging data collected with the central database. For the website part, Windows XP was used as platform, Apache as web server, Microsoft SQL Server as database server, PHP as back-end scripting language and the jQuery framework for implementing user interface features.

## Case Study: Surveying Communication Needs of Busy Parents

ESRM was applied in an investigation of intra-family communication needs and the way pervasive computing would be able to support them to have awareness of each other through the day. This study was aimed at investigating the potential benefits but also the potential disadvantages and detriments to user acceptance of technologies that rely on context sensing to provide frequent updates of the activities and whereabouts of a person to her family see for example Markopoulos et al. (2005). The target group for this investigation was that of "busy parents" who are considered here as pairs of individuals who are married or cohabiting, both working at least part-time and have at least one young, dependent child.

There are several reasons why a sustained and in situ survey of user attitudes was necessary to understand the issues surrounding the awareness needs between couples. First, communication needs vary dynamically; one might wish to reveal her location only when a certain event occurs, e.g., departing from workplace, but not the rest of the day. The same holds for the recipient's interest in such information; it only becomes relevant at particular times and in relation to specific activities. This issue was highlighted during an interview study with 20 busy parents (Khan et al., 2006), a field study of a simple system for helping parents be aware of their children's whereabouts while at school (Khan et al., 2007a) and an online survey with 69 participants (Khan et al., 2007b). These three studies produced results that were sometimes conflicting, only low granularity of information relating to how participants assess experiences for a whole week divorced from its specific time and space context.

Initially it was attempted to answer this question using ESM. However, during pilots with this method involving two participants and two members of the research team for a period of one week, the inherent shortcomings of ESM as discussed above emerged clearly. Pilot study participants perceived the protocol as tedious, which in turn led to repetitive and uninformative answers. There was substantial data loss because of inappropriate timing. Finally, participants missed seeing that their input was actually used and acknowledged by the system. In light of these problems, it was decided to apply the ESRM method using Reconexp as discussed above.

## PARTICIPANTS

Participants fitting the profile of a busy parent were recruited through social relations, through a local community group (scouts) and through a participant database managed by the university. All participants were Dutch citizens, married or cohabiting and were thanked with a gift voucher at the end of the study.

A complication that arose while conducting the study was that some participants had problems synchronizing data. Factors such as firewalls, anti-virus applications or having a proxy prevented 8 out the 20 participants (40%) to synchronize their data and have the opportunity of filling out the unanswered questions posed in the device at the website. Thus the results reported concern 12 out of the 20 participants initially recruited (seven men and five women).

The participants' average age was 38 years (max: 44, min: 28, sd = 5.72); they had on average approximately two children (mean: 1.91, max: 4, min: 1, sd = 0.79) whose average age was 5.47 years (max: 8.5, min: 0.7, sd = 2.57). Participants were married on average approximately 10 years (mean: 10.86, max: 20, min: 2, sd = 5.22) and on average worked close to 30 hours per week (mean: 28.18, max: 40, min: 20, sd = 6.63). Finally, the participant's spouse worked on average also roughly the same hours per week (mean: 30.91, max: 50, min: 20, sd = 8.92).

## PROCEDURE

### Personalization

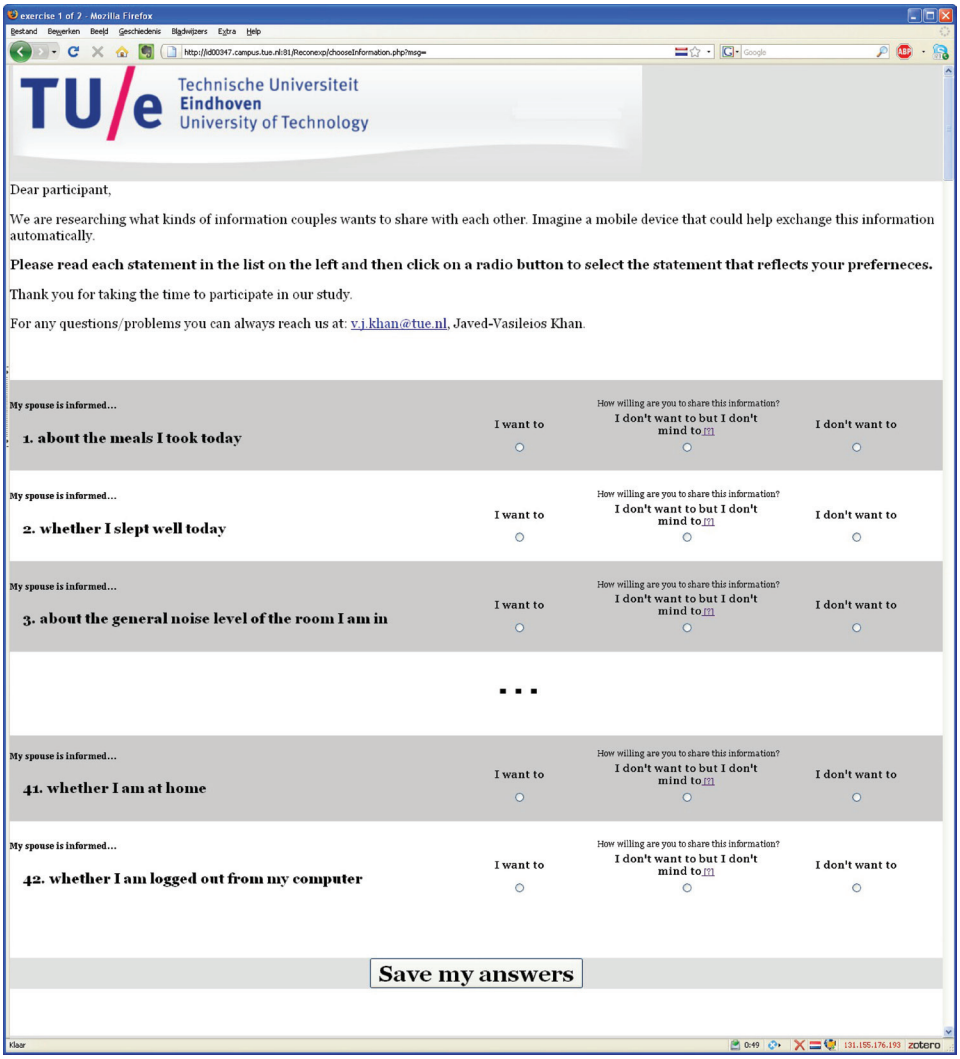
In this section the use of Reconexp for this case study is described. After participants accepted to take part in the study, they were directed to the website where they were initially asked to choose what information they would be interested in sharing with their partner (Figure 2). Rather than a free text entry, participants could select from an extensive list of 41 different types of information to share. This list was compiled from related literature to represent the diversity of information types that is shared through research prototypes described in related literature as well as those mentioned during the interview study mentioned above (Khan et al., 2006). Statements that participants did not wish to share during the personalization step were removed from the options offered during the experience sampling part of the study.

After completing the first part participants were asked to provide information about their context by using the website (Figure 3). The term "context" refers to places and activities (Figure 4) participants visit and perform during a usual working day of theirs. In the final part of this boot-strapping phase on the website participants were asked to imagine what information they would like to communicate while being in a specific place doing a certain activity (Figure 5). This list of information statements was carried from the first part of the personalization step including only those statements that they want or don't mind to share. Participants could add information statements to the list in case they would think of other information that they wanted to share with their spouse in a specific context (Figure 5). Added statements would be presented in the list for all available contexts. This final step was left optional for participants since it would need to be repeated for every place and activity. If for example a participant would have named five places and for each place named five activities then this participant would need to repeat 25 times the last step. That is why participants were allowed to choose the most important contexts and link information statements to those only. In any case, those links would be created during the ESM part.

After a participant inserted the initial information on the website a device was synchronized with the provided data and given to this participant for one week. Participants were requested to keep the device in close proximity constantly.

An audio notification alerted the participant when it was time to record information. Then the application (Figure 6) gave the participant five minutes to respond to three questions: about the

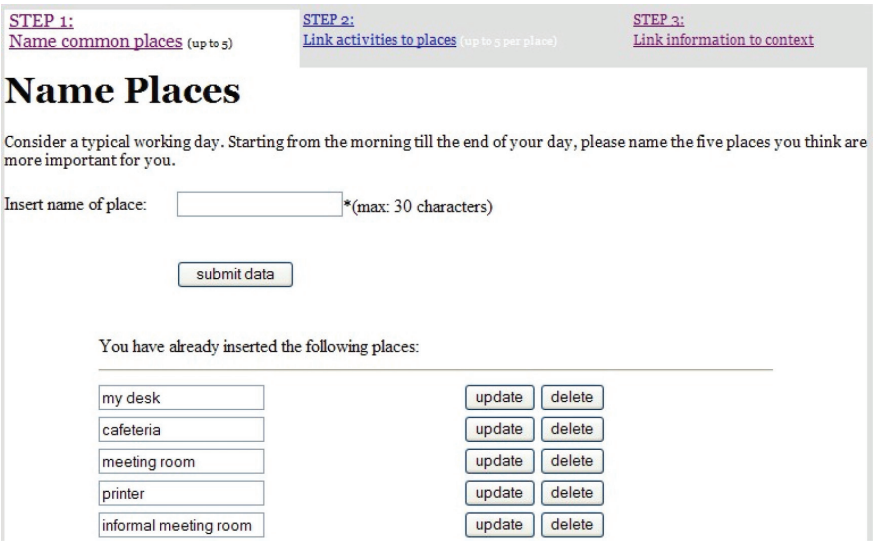
*Figure 2. First part of personalization step: Participants choose information statements they would be generally (in any context) willing to share with their partner, do not mind sharing, or do not want to share*




place where the participant is at the moment, about the participant's current activity and about the information the participant would like to exchange automatically with his/her partner.

Activities and information statements were adapted according to the previous answer. For example, if a participant would answer that he was at his office the next question would present him with the activities he had previously indicated to be doing being at the office. For every question presented on the device, the participant had the option to answer "Other". If a participant would choose "Other" then for the next question no possible answers would be short-listed; rather this participant would be presented with a list-box containing all the activities known to the system (regardless of place). For the last question if the participant

*Figure 3. Second part of personalization step: Naming places visited on a typical working day*




did not check any item from the list, this would be recorded as "Nothing", meaning that the participant did not want to send any information to her/his partner at all.

Furthermore, participants were requested to place the device in its cradle at the end of every day in order to synchronize the data. After synchronizing the data participants were instructed to log onto the website to review the data (Figure 7) correcting omissions of the survey data obtained during the day.

There were two kinds of omissions that participants could correct at this stage. Answers as "Other" (if for example the participant had indicated to be in a place not listed in the drop-down menu offered during the experience sampling) and unanswered questions. When a participant repairs an omission, for example names a place which had previously not been identified, then this new place is appended to the list of places she has made during the personalization phase. This new place will also be present on the mobile device when the participant synchronizes the data. Participants were asked to annotate their answers while reviewing them in order to capture the reasoning behind them.

At the end of the week a semi-structured interview was conducted. In this interview participants were asked whether it had been difficult for them to remember unanswered queries, what information they thought they usually wanted to share and tried to address other issues they might have had with the whole study.

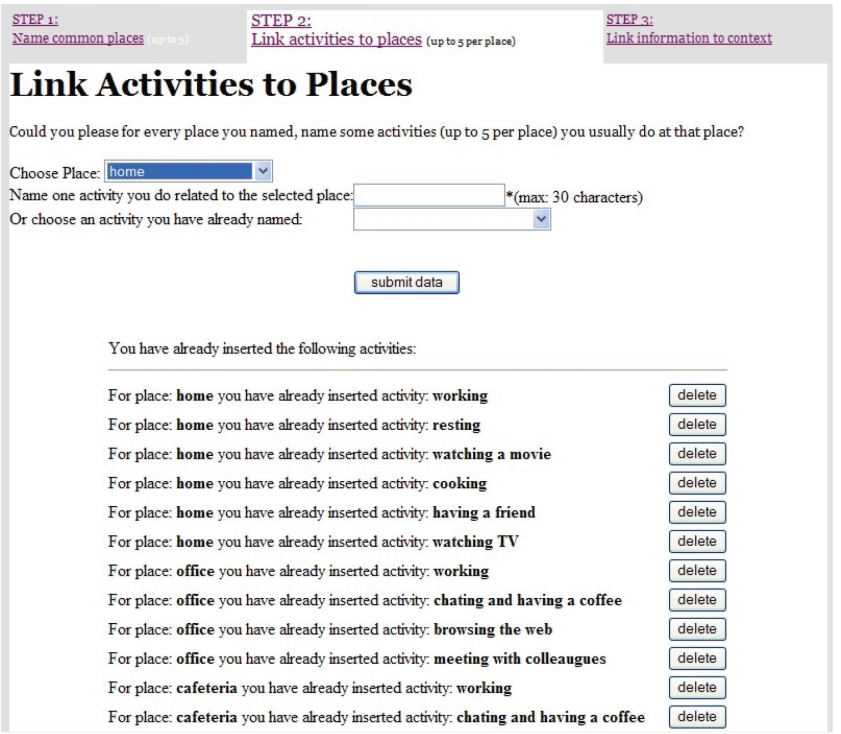
## SAMPLING PROTOCOL

The sampling protocol used combines time based and event based sampling. If an hour has elapsed since the last answer that the participant has given a query is issued. If only 30 to 60 minutes have elapsed the system issues a question only if the location detected by the system has changed. Location changes are detected by comparing the list of Wi-Fi access points to those detected by the system during the most recent answer given by the informant[1].

### Reconstruction Step

Participants were requested to login to a website and review the log of the sampling each day. Participants would then view a log of the

*Figure 4. Third part of personalization step: Naming activities performed on a typical day*




sampling obtained during the day (Figure 7). While reviewing this log, participants could fill out missing queries.

## RESULTS

The twelve participant actions which were logged and their occurred frequency are presented in Table 1.

A first inspection of the data suggests that the website was used heavily. The mean number of actions performed: 55.64 (excluding actions such as logging onto the system).

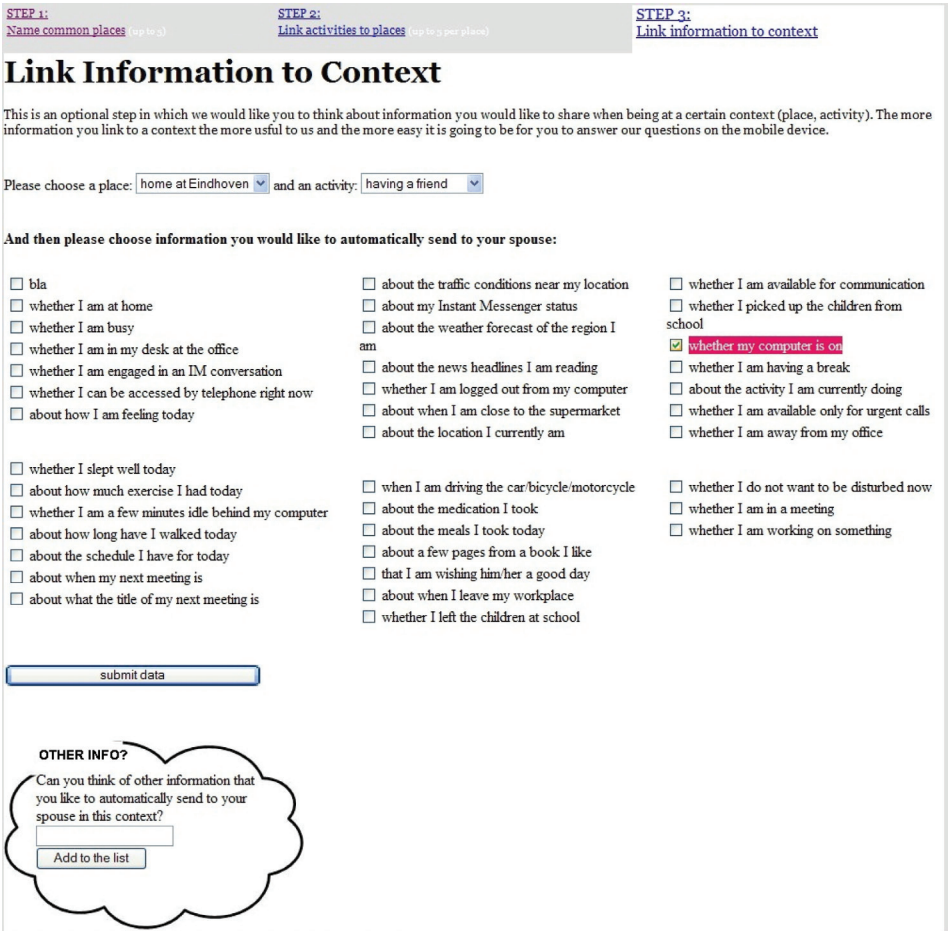
As had been expected, the website was used mostly (64%) during the evening for the 'reconstruction' component of the method.

Furthermore, it turns out that participants logged in 2.75 times in the period of one week to review the data they had inserted during the day. This means that participants would check their log on average approximately once per two sampling days.

In Table 2 the results of the data logging are summarized. It is evident that several questions were not answered during the Experience Sampling component of the method the mean percentage of non response to the second question ("What are you active in now?") was 49.41%. However, a substantial proportion of those (55.38%) were recovered by the use of the website, i.e. via the reconstruction component of the method.

It was about one out of two times that participants could not answer when prompted by the device and it was a just over a half of those that they recovered with the help of the website. When considering answers given for all questions using the device and the website then the overall response improvement of the website to the Experience Sampling Method is: 27.36%. Data pertaining to the second question posed on the

*Figure 5. Fourth part of personalization step: Linking information to context, participants can insert additional information statements to the existing list*

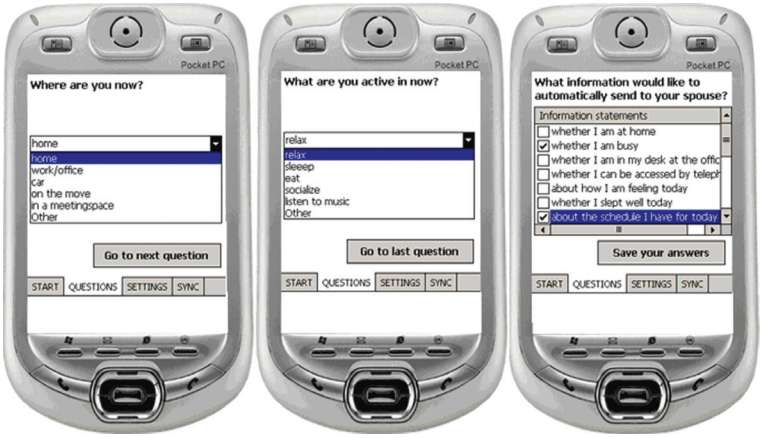



device are analyzed since there was an equal number of times participants were asked the three questions; the recovery of the data for the other two questions differs very slightly.

Places which were named on the website but were never selected during the experience sampling part on the handheld device includes among other: “home”, “doctor”, “train”, “fitness”, “meeting room”, “car”. Activities which were named on the website but were never selected during the experience sampling part on the handheld device include among others: “housekeeping”, “eating”, “cooking”, “shopping”, “check internet”, “put coat on”, “fitness training”, “driving”, “reading news”. Information statements which were named on the website and then never reported using the device were: “dinner time”, “about when he expects to be at home”, “changes in working hours”, “if I can bring something from shop”, “how late I will be home”, “when dinner is ready”, “whether I need to use the car”, “whether the children are going to play elsewhere”, “whether the children are ok”.

The above information illustrates that with the reconstruction phase we were able to obtain reports regarding places and activities where the ESM alone would result in omitted

*Figure 6. The three questions posed at the device (top part of the screens). The drop-down lists in the first two screens and the checkbox list are populated by the personalization step.*




responses. On the one hand this shows that data loss was reduced; on the other it gives us insight into the places and activities in which ESM is less applicable. It is interesting to note that beyond some expected places and activities such as "*car*", "*meeting room*", "*eating*", "*fitness training*" for which we would not have expected participants to be able to respond to the device, we observe places such as "*train*" and "*home*" and activities such as "*check internet*" and "*reading news*" for which participants were not able to respond as well. Thus, the proposed hybrid method helps in extending the scope of ESM as a survey method and its coverage over different contexts.

Furthermore, the content analysis illustrates how reconstruction of places and activities can be extremely precise. In terms of activities for example, the reported activity "*put coat on*" is extremely detailed and it could not be expected that participants could recall such a fine grained activity in a diary study. Both log data and stated opinions seem to corroborate the opposite.

In terms of recalling events we found that the hybrid method does not pose considerable difficulties. During the debriefing interview we asked participants how difficult it was for them to remember the location, activity, and information they would like to exchange when they were asked in the past on the device but could not answer. All but one said that it was easy for them to remember and accurately answer a question they could not answer at the device. They offered two reasons for that. The first one was that when trying to remember and fill out the unanswered questions they concerned situations that were not too remote in the past. In all cases except for one they provided missing information one or two days later.

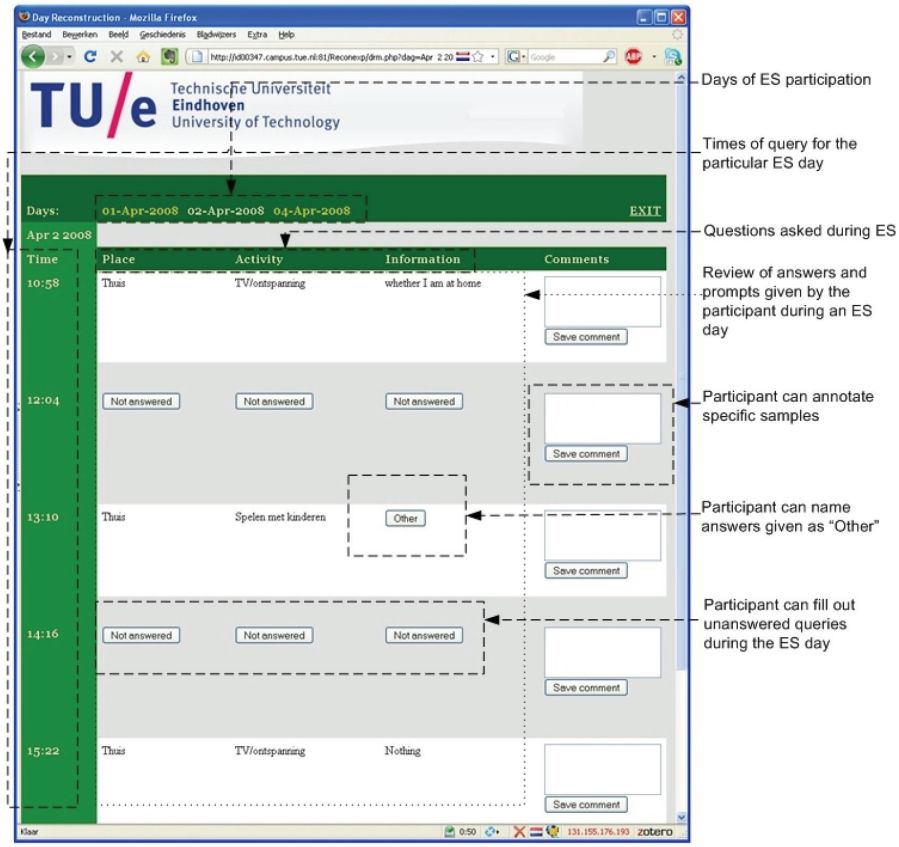
The second reason they mentioned was that the overview on the website was providing them with a frame of reference, (which was the motivation for doing so based on the Day Reconstruction Method). For example, when a participant could not answer a question posed at 13:30 but did answer several questions before and after that, these answers would help recall whereabouts, activities and what information this participant would have liked to exchange in such a context.

## Related Work

In this section we will review current support for computerized Experience Sampling with the aim to identify directions for the development of related tool support in the future.

Notable examples of such tools include: PsycLab Mobile, a tool which supports audio

*Figure 7. The day reconstruction interface*




recording on a pre-defined schedule (Mehl et al., 2001), CAES (Intille et al., 2003), a pioneer in combining sensor input to trigger queries based on events recorded by sensors, Purdue Momentary Assessment Tool (PMAT), which was developed Military Family Research Institute at Purdue University (Weiss et al., 2004) and supports both time based and event based studies and it also provides a desktop application to configure parameters of the ESM study and finally ESP, an open-source tool developed at Boston College (Barrett and Barrett, 2005). The latest version (ESP 4.0) includes a suite of software packages, a Palm OS based application that interacts with participants and a PC application for designing experiments, configuring ESP settings, and collecting data which runs on Windows and Linux. Although these tools have been pioneers in trying to support researchers in conducting ESM they nowadays seem outdated and one of them (CAES) has even discontinued and has joined forces with a more recent tool which will be reviewed in the following paragraphs (MyExperience). Here, more recent studies which have followed and improved upon the early tools are reviewed.

We will begin this review with two projects which have developed tools to integrated aspects of both easily creating ubicomp prototypes as well as evaluating them (Carter et al., 2007, de Sá et al., 2008).

The framework of de Sá et al. (2008) supports both prototyping and evaluation. It provides a log of events taking place on the

*Table 1. Type of participant actions which were logged and their occurrence for the reported study*

| Logged participant action | Occurrence |
|---|---|
| **Login to the system** | 33 |
| **Name a new Location which was not answered** | 11 |
| **Name a Location which was not answered using existing value** | 155 |
| **Name "Other" Location using existing value** | 10 |
| **Name "Other" Location** | 8 |
| **Name a new Activity which was not answered** | 14 |
| **Name Activity which was not answered using existing value** | 155 |
| **Name "Other" Activity using existing value** | 26 |
| **Name "Other" Activity** | 14 |
| **Name Information which was not answered using existing value** | 155 |
| **Name "Other" Information** | 1 |
| **Name "Other" Information using existing value** | 4 |
| **Total participant actions** | **586** |

mobile device and a researcher has also the ability to execute an ESM. It runs in Windows Mobile, Palm OS and Symbian OS devices. The logging engine stores a variety of events. Events range from each tap on the screen, each button press or even each character that was typed by the user. It supports audio and video capture. To analyze the collected data, a log player is provided. The log player resembles a "movie player" which re-enacts every action that took place while the user was interacting with the prototype. ESM can be event-triggered. The framework is publicly available.

A similar open source tool, supporting both mobile prototype creation and remote evaluation of those prototypes is Momento (Carter et al., 2007). It was created to support remote testing of ubicomp applications. Momento can also gather log data, experience sampling data, diary data, and other qualitative data. One of the requirements the researchers found while conducting interviews with ubicomp developers to elicit requirements for their system is the need for integrated tools to allow participants to annotate and review qualitative data. Momento can run on participants' existing networked mobile devices. Researchers can use a desktop application to configure experimental details, to monitor incoming information from participants, send information to participants and review data or export it for further analysis. Momento uses SMS and MMS and HTTP (if available) to share information between participants and researchers. It supports audio, photo and video capture and situated annotation of captured media. The mobile client is configured using a text file. Momento can also support the review and annotation of data by the participants after they have been collected. However it does not support the recovery of data lost during ESM and the developers have not researched the potential benefits of such a feature. A disadvantage is that it needs a desktop installation for both client and server and therefore requires support for troubleshooting and version track management.

MyExperience (Froehlich et al., 2007) is an open source software that supports passive logging of device usage, user context, and environmental sensor readings and active context-triggered user experience sampling to collect in situ, subjective user feedback. Queries can be targeted to moments of interest by triggering sensor readings. Using XML researchers can define survey questions and configure sensors,

*Table 2. Results of logged data*

| Description | | max | min | SD |
|---|---|---|---|---|
| **Mean number of actions performed (logins not counted in this number)** | 46 | 155 | 1 | 64 |
| **Mean logins (in 5 days)** | 2.75 | 7 | 1 | 1.92 |
| **Mean times participants were questioned (at least 5 days)** | 62.58 | 124 | 14 | 27.67 |
| **Mean times the 2nd question (about activity) was not answered** | 30.92 | 65 | 1 | 16.8 |
| **Percentage of mean number of activities (2nd question) not answered:** | 49.41% | | | |
| **Mean of percentage of activities recovered (through the website)** | 55.38% | 97% | 0% | 40% |
| **Overall improvement (data recovery) of the website to the method is** | **27.36%** | | | |
| **Total comments (number of participants who commented: 8 out of 12)** | 33 | 4.13 | 15 | **4.51** |

triggers, and actions. Embedded scripts are used to provide flexibility and expressiveness in specifying the conditions to trigger surveys. MyExperience supports sophisticated survey logic including multiple branching, parameterized questions, and persistent states. It supports audio, photo and video capture. Although it synchronizes collected data opportunistically, it does not provide a web interface for participants to review their answers and fill out the missing queries of the ESM part. While it is designed to run on participant phones and although there extensive work has been done to simplify the installation process it still remains difficult installing it in a Windows Mobile device. Even though it has an incredible amount of features it still requires a long installation process. Moreover, in the case a researcher wants to have real time access it requires to have and manage a MS SQL database server with replication features. Another shortcoming is that it requires removal of the security lock from the phone which might allow programs such as dialers to run without the users' consent.

Another study has revealed that showing participants of experience sampling studies their own collected information helped in increasing compliance rate. In a 25-day field study, Hsieh et al. (2008) found that users who saw visualizations of their own data were more likely to respond to sampling requests compared to users who did not see visualizations. The compliance rate of those receiving feedback was 23% higher than the rate of those who did not receive feedback. This result suggests that showing participants visualizations of their own answers has a positive effect to the study itself. The researchers do not report the use of any tool for conducting the ESM therefore we assume that it was a custom made application.

In their quest to measure and evaluate emotional responses to user interactions with mobile device applications Isomursu et al. (2007) administered an ESM study in which participants answered questions by selecting an appropriate emoticon on the mobile phone's screen. It was a custom made ESM tool which allowed both system and user initiated experience recording. It supported event based triggering of queries. Participants could not insert text or any multimedia input. Further, the tool logged user interactions on the mobile device. The tool was running at the participants' phones. Additional comments could not be provided to clarify the answers. One important finding of their evaluation regarding the method was that participants would accidentally press a button and give an unwanted response in situations like driving, or when having the phone in their coat, or in cases that they would receive a phone call. Such a finding suggests that ESM tools must support defining inconvenient moments

for participants. Having such an option would be beneficial to researchers because they would not collect data which were accidentally inserted and they would not disturb participants with queries during inconvenient moments.

Isomursu et al. (2007) have also created the Experience Clip method. In this method, pairs of participants are recruited. Both of them are provided with mobile phones. One carries the application which is under evaluation and the other is instructed to take short video clips of the usage of the first participant. The participant taking the video clips was a friend of the other participant. Time stamps of the videos were used to match interaction events with the expressions of emotions captured by videos. As a conclusion the researchers state that having the users decide which usage situations to record did not seem to spur versatile and innovative usage. This approach had the disadvantage that it was not clear which captured situations represent real usage situations and which not. They suggest that a combination of the Experience Clips with other non-intrusive methods would perhaps yield better results. This method brings a different perspective into the computerized version of ESM. It advocates the involvement of people in the surrounding of the subject whose experience is sampled.

## CONCLUSIONS AND DIRECTIONS FOR IMPROVEMENT

ESM tools can support behavioral as well as requirements elicitation process for the design and iterative evaluation of mobile applications. Although there are an increasing number of tools available, most of the studies we reviewed still rely on custom-made software to execute ESM studies, which suggests that existing tools do not yet meet the requirements of researchers and practitioners interested to apply the method. Based on the case study findings and the review of existing experience sampling tools a list of directions for future developments can be derived.

From our experience in installing the available tools, we find them still quite challenging to install. It would be even more challenging for a researcher without extensive computer literacy to deploy any of them. All of the reviewed tools require extensive installation procedures and in most cases management of database or web servers. In tools that combine mobile phones and desktop computers, installation procedures have to be followed in both devices. This presents a major obstacle to researchers who lack the relevant technical skills. Quick and easy installations on mobile devices and even no installation procedures for desktops would be another important requirement for such tools. Therefore, special effort should be spent in minimizing requirements for installation and maintenance.

Another important conclusion for future ESM tools is providing an interface for participants to give feedback during the ESM period. Although some of the aforementioned studies have used synchronization of collected data (Froehlich et al., 2007; Hsieh et al., 2008) none of them gave the opportunity to participants to fill out queries they could not answer during the day. This property that was supported by ESRM introduced in this paper, will not only increase compliance rate but it will also recover lost data, as it was presented. Environmental sensor data would also be useful in inferring with greater accuracy the participant's context and in that way make the presentation of queries even more sophisticated at a more convenient time.

The event-triggering of queries, for example when a participant is in a particular location, is supported by some of the tools reviewed however it is either a built-in function, in the case of Reconexp for example, or in the best case (MyExperience) it is programmable by using XML. Although XML is in many ways easier than programming in C or VB it still presents a high threshold for researcher who are not programmers. An appropriate interface for an easily programmable sampling protocol seems an important requirement for the wider uptake of such tools.

The tools reviewed all present the prompts and questions to participants without taking into account whether they are in fact able to observe the screen at that time; some context awareness specifically with regards to the attention of the participant would help optimize the sampling procedure and yield better results.

Support for multimodal participant input has been already included in several tools. Text, audio, photos, and video can provide richer data to the researcher (Carter et al., 2007). On the other hand, participants can choose the most efficient and convenient modality for addressing the query. In the case of CAES, MyExperience, and Reconexp among others, user context factors are automatically captured. The location and possibly the activity of a participant can serve as examples. Automatic capture of participants' context would provide different perspectives for researchers to look at the gathered data and obviously provide more in-depth results.

Another requirement is support for optional, user initiated input. In the case of Reconexp, participants could not initiate the queries at will. That would be useful in cases where participants would recognize the importance, in terms of research, of the context they currently are and initiate the research queries. In that way salient information will be gathered.

An important shortcoming of Reconexp was the difficulty participants had in synchronizing data. Automatic synchronization of data captured on the device to a remote server would both secure the data as well as provide the grounds for feeding the data back to participants as the case with Reconexp was. MyExperience already supports such a feature.

Automatic and configurable information visualization tools of the collected data would be a crucial feature for helping researchers disambiguate the data and quickly provide useful results. Alternative visualizations, e.g. in the form of graphs, can enable researchers to view the data in new fresh ways, and provide insights during analysis. This analysis tool should be able to support visualization of events that occur both frequently and infrequently (Barrett & Barrett, 2001).

Participants might become less motivated during the course of the research study. Programmable by researchers email or SMS notifications to the participants can help to keep participants highly motivated. In addition, support for notifications for researchers when certain events occur would also be of added value.

In most cases, ESM tools which use a mobile device force participants to carry an extra to their own mobile device. It would be more convenient if such tools would run on participants' phones. Such studies could involve larger sample of participants and the reliability of results would be enhanced. However, a researcher would thoroughly need to have tested the tool so that it would not hamper the operation of the participant's device. Moreover, agreements with the mobile service provider must be made in advance so that participants are not burdened with the cost of the service.

Another important feature beyond the ability of participants to review the collected data would be to annotate the data and also to fill-in gaps. The potential benefits of this feature which was implemented in Reconexp have been extensively discussed above.

Mobile devices have limited processing and memory in comparison to desktops. The data collection tool on the mobile device should not noticeably impact the performance of the participant's mobile phone (Carter et al., 2007). If that happens it might affect the results of the study since participants will experience a lag in the presented queries.

Finally, Carter et al. (2007) identify some other important user requirements: In case where a mobile device is lost, the tool on the mobile device should offer mechanisms to protect the security and privacy of the data. Further, to ensure ease of use, participants should be able to increase the color contrast, the font size, etc.

## Discussion

The arguments presented above and the case study show that using Reconexp to support ESRM allows for reduction of data loss and

also, some streamlining of the effort required by participants.

One could raise the question whether the DRM possibly induces "postponing behavior" where participants would opt to postpone answering, exactly because the option is available to them to fill out information easier on the website later. This would mean that the ESM component of the combined method, underperforms, and perhaps accuracy of data is lost as a whole. While there has been no evidence found of such behavior, it is important from a methodological perspective to eliminate that such a behavior undermines the quality of the data obtained. In our future investigations we aim to compare ESM on its own against its adapted version with DRM.

One should note that completion rates with the ESM part of Reconexp (so prior to reconstruction) were considerably lower than those typically reported in the literature. For example, Froehlich et al. (2006) report completion rate of 80.5% and Consolvo and Walker (2003) 80% whereas ours was approximately 52%. This is probably due to the difference in participant groups. The previous studies cite University students as participants. In our case we recruited people who had young children, were working and had a truly hectic schedule. These participants have much less time to respond to an experience sampling study and this is what our results portray. However, further applications of the method are needed to obtain a better understanding of how this set up influences compliance to the amount and quality of the data obtained.

While promising, one needs to keep in mind that Reconexp was a first attempt to support ESRM. There are several usability issues that still need to be tackled and as previously mentioned, there were some issues with synchronizing the data. We hope that ongoing improvements of the functionality and ease of use of the tools should greatly enhance the performance of the method.

## Conclusion and Future Work

This paper has presented a distributed platform (Reconexp – reconstructing experience) developed to support the combination of two research methods for collecting subjective data in field studies regarding experiences and feelings of informants: Experience Sampling and Day Reconstruction. This novel method (ESRM) is part of a line of research to support Experience Sampling tools with the use of mobile devices and context sensing technology.

Compared to related systems, Reconexp is the only one complementing the use of a handheld device for reporting brief notes in situ, with surveying recollection of informants using a website. Related is the work of Froehlich et al. (2007) who have developed a tool which synchronizes data captured during an experience sampling study. Compared to their system Reconexp innovates by allowing participants to review and fill out the gaps created by ESM as well as annotate the captured data.

There is a lot of scope to develop Reconexp further. Our first experiences with Reconexp and ESRM confirm the value of complementing the mobile experience sampling tools with data collection on a website, leading to a reduction of data loss and the improvement of the quality of the data collected. Follow up studies are needed to consolidate these methodological results and to effect relevant improvements on the tools.

## REFERENCES

Barrett, L. F., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, *19*(2), 175–185. doi:10.1177/089443930101900204

Barrett, L. F., & Barrett, D. J. (2005). *ESP: The experience sampling program*. Retrieved from http://www.experience-sampling.org

Carter, S., & Mankoff, J. (2005). When participants do the capturing: The role of media in diary studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 899-908).

Carter, S., Mankoff, J., & Heer, J. (2007). Momento: Support for situated ubicomp experimentation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 125-134).

Consolvo, S., & Walker, M. (2003). Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing / IEEE Computer Society [and] IEEE Communications Society*, *2*(2), 24–31. doi:10.1109/MPRV.2003.1203750

Czikszentimihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease*, *56*, 5–18.

de Sá, M., Carriço, L., Duarte, L., & Reis, T. (2008). A framework for mobile evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2673-2678).

Froehlich, J., Chen, M., Smith, I., & Potter, F. (2006). Voting with your feet: An investigative study of the relationship between place visit behavior and preference. In *Proceedings of the Conference on Ubiquitous Computing* (pp. 333-350).

Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., & Landay, J. A. (2007). MyExperience: A system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services* (pp. 57-70).

Hsieh, G., Li, I., Dey, A., Forlizzi, J., & Hudson, S. E. (2008). Using visualizations to increase compliance in experience sampling. In *Proceedings of the Conference on Ubiquitous Computing* (pp. 164-167).

Hudson, J. M., Christensen, J., Kellogg, W. A., & Erickson, T. (2002). I'd be overwhelmed, but it's just one more thing to do: Availability and interruption in research management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 97-104).

Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., & Bao, L. (2003). A context-aware experience sampling tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 972-973).

Isomursu, M., Tähti, M., Väinämö, S., & Kuutti, K. (2007). Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. *International Journal of Human-Computer Studies*, *65*(4), 404–418. doi:10.1016/j.ijhcs.2006.11.007

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*, 1776. doi:10.1126/science.1103572

Kapoor, A., & Horvitz, E. (2008). Experience sampling for building predictive user models: A comparative study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 657-666).

Khan, V. J., Markopoulos, P., de Ruyter, B., & IJsselsteijn, W. (2007). Expected information needs of parents for pervasive awareness systems. In B. Schiele, A. K. Dey, H. Gellersen, B. de Ruyter, M. Tscheligi, R. Wichert et al. (Eds.), *Proceedings of the European Conference on Ambient Intelligence* (LNCS 4794, pp. 332-339).

Khan, V. J., Markopoulos, P., & Eggen, B. (2007). On the role of awareness systems for supporting parent involvement in young children's schooling. *International Federation for Information Processing*, *241*, 91–101.

Khan, V. J., Markopoulos, P., Mota, S., IJsselsteijn, W., & de Ruyter, B. (2006). Intra-family communication needs: How can awareness systems provide support? In *Proceedings of the 2nd International Conference on Intelligent Environments* (Vol. 2, pp. 89-94).

Markopoulos, P. (2005). Designing ubiquitous computer human interaction: The case of the connected family. In Isomaki, H., Pirhonen, A., Roast, C., & Saariluoma, P. (Eds.), *Future interaction design* (pp. 125–150). New York, NY: Springer. doi:10.1007/1-84628-089-3_8

Mehl, M. R., Pennebaker, J. W., Crow, M. D., Dabbs, J., & Price, J. H. (2001). The electronically activated recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, *33*, 517–523. doi:10.3758/BF03195410

Rieman, J. (1993). The diary study: A workplace-oriented research tool to guide laboratory efforts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 321-326).

Weiss, H. M., Beal, D. J., Lucy, S. L., & MacDermid, S. M. (2004). Constructing EMA studies with PMAT: The Purdue momentary assessment tool user's manual. Retrieved from.http://www.ruf.rice.edu/~dbeal/pmatusermanual.pdf

## ENDNOTE

[1] If in a previously stored place A the system has detected N1 access points and if N2 is the number of access points of N1 which were not found in the surrounding access points at the query moment, the following formula was used to approximate the probability of the user being at A:

$$p = \log \frac{(N_1 - N_2) * 10}{N_1},$$
$$for \frac{(N_1 - N_2) * 10}{N_1} > 1 \wedge N_1 > 0$$

It appears that this formula can easily flag when an informant moves to a different building, floor or part of town based on the current density of wi-fi points. This approach does not help track fine grain movements, e.g., away from and to a desk, for which other approaches may be more appropriate.

*Panos Markopoulos studied electrical engineering and computer science in the National Technical University of Athens and human-computer interaction in Queen Mary University of London, where he did a PhD on formal methods in human computer interaction. He has held research positions in Queen Mary, University of London and Philips Research in Eindhoven and is currently a Professor in the Department of Industrial Design of the Eindhoven University of Technology on the topic of Awareness Systems and Ambient Experiences. His interests extend in several areas of the fields of human-computer interaction and interaction design such as model based design, privacy and trust, social games for children, evaluation methodology for children, connectedness between family and friends, persuasive technologies, and healthcare. Panos Markopoulos has published extensively in all these topics, and is the first co-author of a book on evaluating children's interactive products, published by Morgan Kaufmann in 2008.*

*Vassilis-Javed Khan is a senior lecturer at NHTV Breda University of Applied Sciences, in the Netherlands. He completed his PhD and P.D. Eng. in Eindhoven University of Technology in the Netherlands and his MSc from the University of Patras in Greece. He has worked as a post-doctoral researcher at Eindhoven University of Technology, visiting researcher at Philips Research in Eindhoven, research intern at Vodafone R&D in Maastricht and as a software engineer at the Computer Technology Institute in Patras, Greece. His research interests fall under the umbrella of HCI and include pervasive computing, location based services, awareness systems and game evaluation. An entrepreneurial side of his research is KidzFrame which he has co-founded.*